

NBER WORKING PAPER SERIES

THE TRANSFER OF HUMAN CAPITAL

Boyan Jovanovic
Yaw Nyarko

Working Paper No. 4823

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1994

We thank the C.V. Starr Center for Applied Economics at New York University for technical and financial help, and Jacob Mincer and Nachum Sicherman for useful comments, and Chung Tse for help with the research. This paper is part of NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1994 by Boyan Jovanovic and Yaw Nyarko. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

THE TRANSFER OF HUMAN CAPITAL

ABSTRACT

Most of our productive knowledge was handed down to us by previous generations. The transfer of knowledge from the old to the young is therefore a cornerstone of productivity growth. We study this process in a model in which the old sell knowledge to the young -- old workers train the young, and charge them for this service. We take an information-theoretic approach in which training occurs if a young agent watches an old worker perform a task. This assumption has plenty of empirical support -- in their first three months on the job, young workers spend about five times as long watching others work as they do in formal training programs.

Equilibrium is not constrained Pareto optimal. The old have private information, and this gives rise to an adverse selection problem: some old agents manage to sell skills that the young would not buy (if only they knew exactly what they were buying). We derive the implications for the lifetime of technological lines, and we show that the model generates a negative relation between a firm's productivity and its probability of failure.

Boyan Jovanovic
Department of Economics
New York University
269 Mercer Street
New York, NY 10003
and NBER

Yaw Nyarko
Department of Economics
New York University
269 Mercer Street
New York, NY 10003

1. Introduction

How efficiently is productive knowledge passed on from one generation to the next? Some of this knowledge is handed down informally in the family, and some of it is taught in schools. A third mechanism is learning on the job: acquiring productive knowledge through the process of on-the-job training and learning by doing. Our paper focuses on the third mechanism.

Two prominent equilibrium models of learning on the job are those of Prescott and Boyd (1987), and Chari and Hopenhayn (1991). They assume that the process by which one worker learns from other workers can be (and is) fully internalized by firms. Our approach is similar in spirit, but our conclusions differ from theirs. We explicitly model the information that the process of training transfers, and this leads to the conclusion that the market for on-the-job training is plagued by the efficiency problem that usually crops up with markets for information.

We study an overlapping generations economy in which people live for two periods. An old worker trains the young worker who works alongside him. This training consists of an explicit mechanism by which information is revealed to the young worker. The information consists of observations of past signals about the unknown parameters of the production process that the old and the young worker alike both use. So although agents die, their information can survive for ever.

There are two fundamental differences between this economy and a frictionless one in which a pair of infinitely lived workers were to share their information instantaneously; both of them imply a level of output that is lower in the overlapping generations economy than in the infinite horizon one. The first difference occurs because the transfer of knowledge is assumed to take time: the young worker must watch the old worker do the job¹. Although the young worker can work and watch at the same time (so that training is assumed to involve no foregone output cost), information still accumulates more slowly than it would if an infinitely lived agents were to observe the same signals.

¹ The assumption that the transfer of knowledge occurs through watching is backed by a study by Barron, Black and Loewenstein (1989) who report (in table 1) that in a sample of some 2,000 employers, the typical worker spent at least five times as many hours watching others work as he did in formal training programs.

This is because there is no way for the old worker to pass on the information that he gains from the signal that he sees in the second period of his life. Therefore, in the overlapping generations economy, the two-worker team will have only half the productive information that a pair of infinitely lived individuals would have, and a permanently lower level of output around trend. This first difference therefore simply reflects the added constraints that we impose on informational transfer among people.

The second difference stems from a failure in the market for information. We follow Prescott, Boyd, Chari, and Hopenhayn assume that the old sell training to the young. The content of the training is known to the old worker, but not to a prospective young buyer. The young worker therefore enters a match with an old worker based on what he knows about the population of all old workers. Old workers with inferior skills exploit this, and in equilibrium too much training occurs in comparison to what a central planner could achieve through a tax scheme. Equilibrium allows the survival of some skills that the planner would prefer to extinguish so as to release more young agents to start new technologies. In this sense, competitive equilibrium involves **too much resistance to change**.

2. Model

There will be many technological lines. We first describe what happens within one such line.

2.1. Within a Technological "Line:" A technological line has many grades indexed by $n=0,1,2,\dots$. The production function and information structure within a technological line are taken from Jovanovic and Nyarko (1994). If an agent uses grade n at date t , his decision z yields net output q via the production function

$$q = \gamma^n [1 - (y_n - z)^2]. \quad (1)$$

Here y_n is a random variable that acts as an unknown "target", and is observed after z is chosen.

Since $\gamma \geq 1$, a larger n denotes a better grade. Since y_{nt} is observed regardless of what z is chosen, z does not affect learning. Assume that everyone is risk-neutral, so that each agent maximizes his own expected output. Let $E_t(\cdot)$ denote the conditional expectation at date t . The decision that maximizes $E_t(q)$ in (1) is

$$z = E_t(y_{nt}) . \quad (2)$$

The random target fluctuates around a grade-specific parameter θ_n :

$$y_{nt} = \theta_n + w_{nt} . \quad (3)$$

The agent does not know θ_n . He can observe y_{nt} but only if at date t he uses grade n . Let $N(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2 . Assume that w_{nt} is an i.i.d variate, with marginal distribution $N(0, \sigma_w^2)$. Since $E_t(w_{nt}) = 0$, equation (2) implies that the optimal decision is

$$z = E_t(\theta_n) , \quad (4)$$

and (1) - (4) imply that expected net output is

$$E_t(q) = \gamma^n [1 - \text{Var}_t(\theta_n) - \sigma_w^2] , \quad (5)$$

where $\text{Var}_t(\cdot)$ denotes the conditional variance. If he uses grade n , he observes y_{nt} and learns more about θ_n , which allows him to make a better decision z . This reduces $\text{Var}_t(\theta_n)$, and raises his expected net output. Indeed, suppose that the agent has a prior belief over θ_n which is $N(m_n, x_n)$ where

$$m_n = E_t \theta_n \quad \text{and} \quad x_n = \text{var}_t \theta_n . \quad (6)$$

Then upon observing y_n it is easy to verify that the updated beliefs over θ_n will be normal with variance given by the function

$$h_1(x) \equiv \sigma_w^2 x / (\sigma_w^2 + x); \quad (7)$$

in particular the posterior variance is $x_n^* = h_1(x_n)$. The updated posterior mean is given by

$$m_n^* = h_1(x_n) \{m_n/x_n + y_n/\sigma_w^2\}. \quad (8)$$

However this learning process is bounded: Using grade n forever allows the agent to learn θ_n completely so that $E_i(q) \rightarrow \gamma^n [1 - \sigma_w^2] < \infty$.

2.2. Upgrading Within a Technological Line: Successive grades are linked as follows:

$$\theta_{n+1} = \sqrt{\alpha} \theta_n + \epsilon_{n+1}, \quad n = 1, \dots \quad (9)$$

The ϵ 's are iid with distribution $N(0, \sigma_\epsilon^2)$. ϵ_{n+1} and θ_n are independent of each other. Agents know α and σ_ϵ^2 , but they do not see the θ 's or ϵ 's.

If x_n^* is an agent's posterior variance over θ_n at the end of the period, then from (9) the agent's posterior variance, x_{n+1}^* , over θ_{n+1} , will be given by the function h_2 below of x_n^* :

$$h_2(x) = \alpha x + \sigma_\epsilon^2. \quad (10)$$

Hence that $x_{n+1}^* = h_2(x_n^*)$. Suppose that at the beginning of the period an agent has a posterior variance over θ_n equal to x_n . Suppose that agent uses grade n . Then from (7) and (10) we

conclude that the mapping from x_n into x_{n+1} is given by the mapping

$$B(x) = h_2(h_1(x)) , \quad (11)$$

so that $x_{n+1} = B(x_n)$. If m_n is the mean of the beliefs over θ_n , then from (8) and (9), after y_{nt} has been observed the posterior mean over θ_{n+1} will be

$$m_{n+1} = \alpha^{1/2} h_1(x_n) \{ m_n / x_n + y_{nt} / \sigma_w^2 \} . \quad (12)$$

It is easily verified that B has a fixed point which we denote by \bar{x} . (See fig. 1). We assume that $|\alpha| < 1$. The map on the right-hand side of eq. (10) then has a fixed point

$$\hat{x} \equiv \sigma_e^2 / (1 - \alpha) . \quad (13)$$

From \hat{x} iterates of the function B converge monotonically to \bar{x} . (See fig. 1).

[Figure 1 about here]

We suppose that the ex ante distribution over θ_0 , the parameter of grade 0, is its stationary distribution - $N(0, \hat{x})$. We assume that

$$1 - \hat{x} - \sigma_w^2 > 0. \quad (14)$$

This means that the ex ante expected output on grade zero is positive.

2.3. Switching to a New Technological Line: We suppose that there is a continuum of technological lines available. Ex ante, each technological line is the same. Each technological line has an associated sequence of grades $n=0,1,2,\dots$, and parameters $\{\theta_n\}_{n=1}^\infty$ as described in sections 2.1

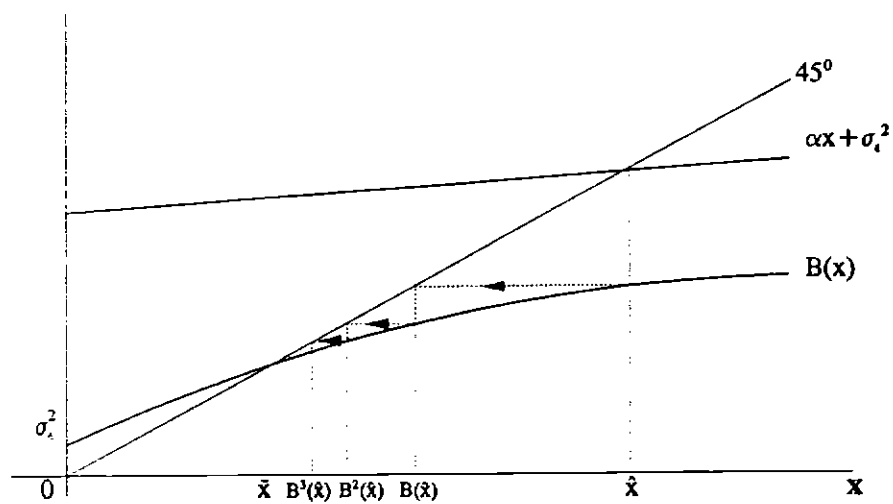


Figure 1: The Determination of \bar{x} , \bar{x} and $B^1(x)$.

and 2.2. The $\{\theta_n\}_{n=1}^{\infty}$ sequence associated with one technological line is independent of the sequence associated with any other technological line. We impose the assumption

(A1) On an untried technological line, the ex ante belief over θ_0 (and hence over θ_n for each $n=1,2,3,\dots$) is $N(0,\hat{\lambda})$.

At date t , any agent can choose on any technological line any grade n which is less than or equal to t . This means that the "production frontier" is moving out in step with time. (In Jovanovic and Nyarko (1994) this was called the Chari-Hopenhayn assumption.)

Suppose the agent is contemplating using a technological line over which he has no information. From (A1), his ex ante variance over θ_n is equal to $\hat{\lambda}$ for all n . Then (14) implies that his expected output is the highest if he sets $n=t$. If an agent abandons a technological line whose grade is on the frontier, and starts a new technological line, the switch will represent a *lateral* switch - i.e., to the same grade $n=t$. (Jovanovic and Nyarko (1994) had only one technological line. Lateral switches were therefore not allowed in that paper.) In the model we present in the next section, some agents will indeed make lateral switches.

3. The Overlapping Generations Structure

We shall assume an overlapping generations structure with two-period lives. People are risk-neutral, with lifetime preferences $c_{\text{young}} + \delta c_{\text{old}}$ where c_{young} and c_{old} denote the consumption when young and when old respectively, and where δ is the discount factor which we assume satisfies $0 < \delta < 1/\gamma$.

If the young and the old work together, the young can learn from the old. We shall refer to an old agent as "the foreman". A young agent working with a foreman will be referred to as "the apprentice." All decisions are made at the beginning of a period. Anything learned during the period can not affect decisions until the following period.

We shall make the following assumption:

- (A2) Agents are always on the production frontier (in the sense of $n=t$) of whatever technological line they are using.

We will provide justification for assumption (A2) in Appendix B.

3.1. The Foreman's Problem: First consider an old agent - a foreman - at some date t . Suppose that the foreman is using a technological line that has been in use since date 0. By assumption the foreman will be at the frontier, and hence will be using a grade $n=t$. Therefore n is also the age of the technology. Suppose that the belief of the foreman over the parameter θ_n is $N(m, x)$. The foreman will operate that grade according to (4) and will receive an expected payoff equal to $\gamma^n(1 - \sigma_w^2 - x)$. The foreman at the beginning of the period also has to decide on whether or not to hire a young agent as an apprentice. If the foreman does not hire an apprentice he receives no further payoffs. He dies. So does that technological line!

Suppose on the other hand that the foreman does hire an apprentice. The apprentice will then use the technology of the foreman, and his grade, to produce output. As assumed in section 2.3, the apprentice, when hired, has a belief over θ_n equal to its stationary distribution, $N(0, \hat{x})$. With this belief the apprentice chooses an action according to (4). In particular, he chooses the action $z=0$. From (1) this results in an output of $1-y^2$. We suppose all of this output goes to the foreman. We now state the outputs and payments in units deflated by γ^n . Using (3), the expected value of this output in terms of the expectations operator of the foreman, E_t , is given by:

$$E_t(1-y^2) = 1 - \sigma_w^2 - x - m^2. \quad (15)$$

In return for these labor services, the foreman pays the apprentice an amount c_n (which we shall

characterize later). We refer to c_n as the wage rate associated with labor on grade n . We may represent the value function and decision problem of the foreman by

$$V_n(m, x) \equiv \text{Max} \{1 - x - \sigma_w^2, 1 - x - \sigma_w^2 + E_f(1 - y^2) - c_n\}. \quad (16)$$

Equation (16) indicates that the decision of the firm is to hire or not hire. The definition in (16) is stated as if n and x are both free variables. We will later only study values of x which are some deterministic function of n , in which case we drop the x from the specification of the value function.

3.2. The Young Agent's Problem: The young may either be hired by a foreman or not hired. We first consider the case of an agent who is hired.

3.2a. If agent is hired: We now consider the situation of a young agent who is hired by a foreman using a grade n of a technological line. The two, the young and the old that is, must then use the same grade. The apprentice will receive the amount c_n in his first period as payment from the foreman. At the end of the period the young will receive information from two sources. The first of course is the observation of y_n from use of that grade. The apprentice also learns by watching: he obtains information from observing the foreman's decision, z . The foreman sets z to equal his own expectation of θ_n , $E_t\theta_n$. Having seen his z exactly, the apprentice learns the foreman's $m_n = E_t(\theta_n)$. Each foreman has a reputation in that it's widely known how much he knows, but not exactly what he knows. That is, the foreman's posterior variance over θ_n will be common knowledge. We shall denote this posterior variance by x_n . Since the foreman's beliefs are normally distributed, this implies that his beliefs are fully revealed once his z is observed.

These two sources of information leaves the apprentice with beliefs over θ_{n+1} , the parameter of the next grade, which, from (9) is easily seen to be normal with mean and variance given by (12) and (11). In the next period the young agent becomes old. He becomes a foreman and solves

the foreman-problem described earlier, but with grade $n+1$. His lifetime return (deflated by γ^n) is therefore

$$c_n + \delta \gamma E_a[V_{n+1}(m_{n+1}, x_{n+1})] , \quad (17)$$

where E_a denotes his expectations operator over m_{n+1} when he is young. We shall define E_a precisely in equation (30).

We assume that it takes an entire period for the young to observe the action choice of the old. In particular, the old can not merely inform the young of his beliefs at the beginning of the period before production takes place. We envisage a situation where the young actually have to watch the old, and that this takes an entire period. Each apprentice therefore chooses an action $z=0$. The foreman with beliefs which are normal with mean m however believes that the optimal decision is $z=m$. Hence that foreman believes that the young will make "mistakes." When m is large those mistakes are large. Hence not every foreman will take on an apprentice. Foremen will have skills specific to technologies on which young agents are highly unproductive. This would be a technology on which $|\theta_{||}$ has revealed itself (to the old agent) to be large. If hired, a young agent make costly mistakes. Such technologies will therefore die with the foreman who operates them, and new technologies will be started by young agents that did not join up with a foreman.

3.2b. *"Going it Alone:"* Suppose that the young agent is not hired by the foreman who was about to use grade n of a technological line. The young agent then starts a new technological line. From (A2) the young agent uses grade n of the new technological line. From (A1), the young agent's belief over the parameter of grade n of the new technological line is $N(0, \hat{x})$. This is of course the same as the ex ante distribution of θ_0 of the original technological line. Indeed, after deflating by γ^n , the expected output of this agent on the new technological line is equal to the expected output on grade 0 of the original technological line. This deflated output is equal to $1 - \hat{x} - \sigma_w^2$. Let y_0 denote his observation from the use of that grade. Then from (12) and (11) his posterior belief over the

next grade after observing y_0 is $N(m_1, x_1)$ where

$$m_1 = \alpha^{1/2} h_1(\hat{x}) y_0 / \sigma_w^2 \quad \text{and} \quad x_1 = B(\hat{x}). \quad (18)$$

His expected lifetime utility, discounted by γ^n , is therefore given by

$$u_0 \equiv 1 - \hat{x} - \sigma_w^2 + \delta \gamma E_x V_1(m_1, x_1), \quad (19)$$

where E_x denotes expectations over the random variable m_1 defined in (18). u_0 is the value to the young of "going it alone" - when not hired by a foreman. This is in units discounted by γ^n . Implicit in (19) is a stationarity assumption which says that calendar time matters only in as much as it affects production possibilities. We say more about this in section 5.

3.2c. Young agents are perfect substitutes. Old agents are not; they have learned a specific skill. In particular they have information on a technological line. For now we assume that all the rents from the partnership go to the old. Later on we show that some young agents are not hired in equilibrium. Competition then results in the old receiving all the rents. This implies that in equilibrium the value to the young from being hired, i.e., (17), will equal the return from going it alone, (19). This implies that

$$c_n = 1 - \hat{x} - \sigma_w^2 + \delta \gamma \{E V_1(m_1, x_1) - E_x[V_{n+1}(m_{n+1}, B(x_n))]\}. \quad (20)$$

Putting this value of c_n into the foreman's problem (16) and simplifying implies that

$$V_n(m, x) = 1 - x - \sigma_w^2 + \text{Max} \{0, g_n - m^2\}, \quad (21)$$

where g_n is independent of m and is given by

$$g_n = \hat{x} - x_n + \delta \gamma [E_2 V_{n+1}(m', B(x_n))] - E_2 V_1(m_1, B(\hat{x}))]. \quad (22)$$

Suppose that the foreman is of type (m, x) with $m=0$. Then that foreman would himself set $z=0$ when operating his technology. A young agent using his technology always sets $m=0$. The expression g_n above is the net output such a foreman receives from the young, $E(1-y^2) - c_n$. If the old is of type (m, x) with $m \neq 0$, that old person would prefer the young to choose the decision $z=m$. However, the young always choose action $z=0$. As far as the old is concerned, the young would be making "mistakes" when operating his technology. The total net output that the old receive from the young if hired, is $g_n - m^2$. The amount $-m^2$ therefore represents the losses or costs due to the mistakes of the young. If these are large the old will then decide not to hire the agent.

A foreman of type (m, x) will therefore hire a young agent at date n if $g_n - m^2 > 0$ (and will be indifferent between hiring and not hiring if equality holds). We define

$$\begin{aligned} \bar{m}_n(x) &\equiv (g_n)^{1/2} \text{ if } g_n > 0, & \text{and} \\ \bar{m}_n(x) &\equiv 0 \text{ otherwise.} \end{aligned} \quad (23)$$

A foreman of type (m, x) will hire a young agent whenever m lies in the set $A_n(x) \equiv [-\bar{m}_n(x), \bar{m}_n(x)]$. We call $\bar{m}_n(x)$ a hiring rule, and $A_n(x)$ a hiring set. A foreman is more likely to hire a young agent the smaller is the foreman's $|m|$.

The $A(x)$ correspondence defines the hiring set in (m, x) space, and it is shown as the shaded area in figure 2. Recall that \bar{x} is the fixed point of the operator B . Since $B(\hat{x}) < \hat{x}$, and $B(x)$ is as shown in figure 1, the sequence $B^i(\hat{x})$ declines monotonically from \hat{x} to \bar{x} as shown in this figure. This is why in figure 2, values of x below \bar{x} or above \hat{x} could not arise.

[figure 2 about here].

At the date 0 we suppose that the old at that date have the same beliefs about the technology as do the young. In particular, $m_0 = 0$ and the variance of m_0 is zero. The young receive no extra

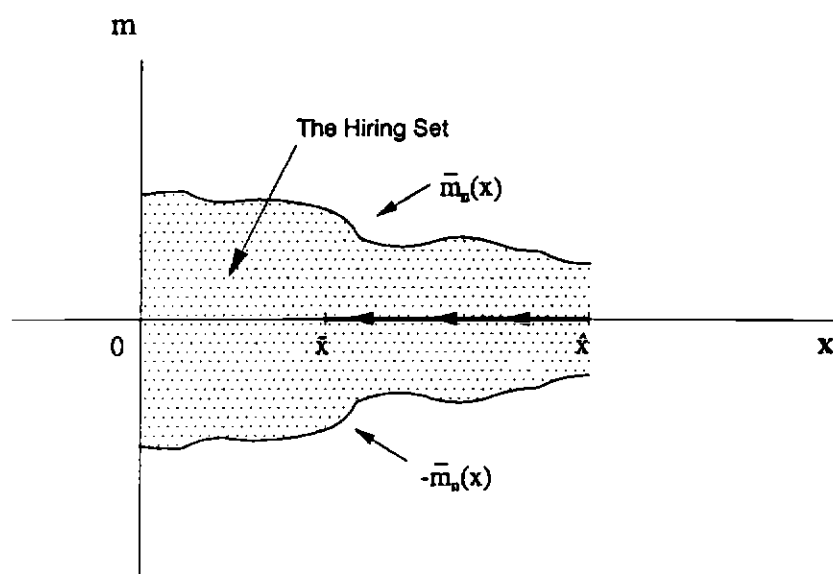


Figure 2: The Hiring Set.

information after using the technology of the old at that date as opposed to going it alone. If indeed the young were hired they would be paid an amount when young, c_n , equal to what they would receive from going it alone $1 - \sigma_w^2 - \hat{x}$.

3.3. The Transfer of Information Along a Technological Line First consider the following slight digression. Suppose two agents A and B have the *same* prior over some parameter θ . Suppose A receives some information and forms a posterior over θ . Suppose that B does not receive that information but observes the posterior of agent A. What then is B's posterior over θ ? It will of course be equal to the posterior of A. Despite not knowing precisely the information of A, B says to himself "any information that A *could have* seen would lead me to have a posterior equal to the announced posterior of A since I have the same prior as A. Hence, despite the fact that I did not see A's information, my posterior is the same as that of A."

The agent who is young at date 0 has a prior $N(0, \hat{x})$ over θ_0 . That agent then observes y_0 and uses this to form a posterior over θ_0 , and hence over θ_1 . His posterior over θ_1 is then $N(m_1, x_1)$. Suppose he hires at date 1 (or more precisely the first period in the life of the technological line) and announces m_1 to the young at date 1. The young at that date do *not* see y_1 ; only m_1 . What will be that agent's belief over θ_1 ? From the argument made earlier, the young at date 1 will have the same posterior, $N(m_1, x_1)$, over θ_1 as the old at date 1. In particular, observation of m_1 is equivalent to observation of y_0 as regards forming a posterior over θ_1 . We may extend this argument inductively to each period n . In particular, consider the young who observe m_n , the posterior mean over θ_n of the old. Conditional on observing m_n , the young will have a posterior over θ_n which is $N(m_n, x_n)$. In particular, the young will inherit the beliefs of the old. Furthermore, the belief of the young conditional on observing m_n is the same as the beliefs that the young would have if they had observed all of the signals $Y_{n-1} \equiv \{y_0, y_1, \dots, y_{n-1}\}$ that all agents of earlier periods had observed on that technological line. m_n is a sufficient statistic for the entire information vector Y_{n-1} .

A foreman using a grade n will therefore have a posterior variance over the parameter θ_n

equal to $x_n = B^a(\hat{x})$. The value function of the foreman is therefore a function only of his prior mean m_n and the age of the technology. We define

$$V_n(m) \equiv V_n(m, B^a(\hat{x})) \quad , \quad \bar{m}_n \equiv \bar{m}_n(B^a(\hat{x})) \quad , \quad \text{and} \quad A_n \equiv A_n(B^a(\hat{x})) \quad . \quad (24)$$

and use this new notation throughout.

4. The $\{m_n\}_{n=1}^{\infty}$ Process

4.1. Suppose that the posterior belief over θ_n is $N(m_n, x_n)$. From (12) and (3), after the signal y_n has been observed m_{n+1} , the posterior mean over θ_{n+1} , is given by

$$m_{n+1} = \alpha^{1/2} h_1(x_n) \{m_n/x_n + ((\theta_n + w_n)/\sigma_w^2)\} . \quad (25)$$

Conditional on m_n , the posterior over θ_n is $N(m_n, x_n)$. Hence conditional on m_n , m_{n+1} is normal with mean and variance given below:

$$E[m_{n+1}|m_n] = \alpha^{1/2} h_1(x_n) m_n \{1/x_n + 1/\sigma_w^2\} = \alpha^{1/2} m_n ; \quad (26)$$

$$\text{Var} [m_{n+1}|m_n] = [\alpha^{1/2} h_1(x_n)]^2 (1/\sigma_w^2)^2 [x_n + \sigma_w^2] = \alpha h_1(x_n) (1/\sigma_w^2) x_n ; \text{ and} \quad (27)$$

$$E[m_{n+1}^2|m_n] = \alpha h_1(x_n) (1/\sigma_w^2) x_n + \alpha m_n^2 . \quad (28)$$

Equations (26) and (27) completely describe the distribution of m_{n+1} conditional on m_n , which we denote by $\text{Prob}(\cdot|m_n)$. The unconditional variance of m_n is given by the recursion

$$\text{Var } m_{n+1} = \alpha h_1(x_n) (1/\sigma_w^2) x_n + \alpha \text{Var } m_n \quad \text{and} \quad \text{Var } m_0 = 0. \quad (29)$$

We define

$$E_s = E[. | A_1, \dots, A_n]. \quad (30)$$

to be the expectation operator over the $\{m_n\}_n$ process conditional on some hiring sets A_1, \dots, A_n , where $A_r = [-\bar{m}_r, \bar{m}_r]$ for some $\bar{m}_r > 0$. (This operator was introduced in (17).)

We now have the following:

Proposition 4.1: Fix any n .

- (a) Prob $[m_{n+1}^2 \leq m_0 | m_n = m]$ decreases as $|m|$ increases.
- (b) For each n , Prob $[m_{n+1}^2 \leq m_0 | m_n \leq m]$ decreases as $|m|$ increases.
- (c) The distribution of m_{n+1}^2 (resp. $|m_{n+1}|$) under the probability defining the conditional expectations $E[. | m_n = \bar{m}_n]$ first order stochastically dominates distribution of m_{n+1}^2 (resp. $|m_{n+1}|$) under the probability defining E_s in (30) where $A_r = [-\bar{m}_r, \bar{m}_r]$.

Corollary 4.2.: The distribution of any increasing monotone transform of m_{n+1}^2 (e.g., $|m_{n+1}|$) is also similarly stochastically increasing in $|m_n|$. In particular all statements in Proposition 1 also hold when m_{n+1}^2 is replaced with an increasing monotone transform of m_{n+1}^2 .

Remark: Fix m_0 and m . Prob $(m_{n+1}^2 \leq m_0 | m_n = m)$ can not be similarly ordered in terms of n .

4.2. The young Agent's Ex ante belief about m_n . Consider a young agent who is considering whether or not to work with a foreman who has a technology of vintage n . What will be that agent's belief about the value of the posterior mean of the old agent, m_n . Note that this belief and

(12) determines the young agents expectation operator E_n over m_{n+1} which is what was required in (17) to state the return to the young from working with the old.

The young will not use the unconditional distribution, determined by (26) and (27) above to form expectations over m_n . The young agent knows the age, n , of the technology. This implies that the foremen in previous periods all hired young agents. This in turn implies that the m_t process fell in the hiring set A_t in each of the previous periods of the life of that technological line. The young agent will condition on this information in forming a beliefs about the value of m_n .

Fix a sequence of hiring sets $A_n = [-\bar{m}_n, \bar{m}_n]$. In particular, an old agent using a technology of age n will hire a young agent if and only if his mean m_n lies in the set A_n . Let σ_n^2 be the variance of m_n conditional on knowledge that m_t was in A_t for $t=0, \dots, n-1$. This is the belief of an agent whose only information is that the technological line has been alive in periods $t=0, 1, \dots, n-1$. Then we have

Proposition 4.3: σ_{n-1}^2 obeys the pair of inequalities

$$\frac{\alpha h_1(x_n)x_n}{\sigma_w^2} \leq \sigma_{n-1}^2 \leq \frac{\alpha h_1(x_n)x_n}{\sigma_w^2} + \alpha \sigma_n^2,$$

with $\sigma_0^2 = 0$.

5. Existence and Properties of Equilibrium: We now define an equilibrium for the model. Our definition will implicitly require that some amount of hiring takes place on each grade n , at least for the foreman with $m_n=0$. We show in appendix A that any equilibrium must be of this kind.

Definition: An equilibrium is a sequence of wage rates $\{c_n\}_{n=0}^{\infty}$ such that if we define

$$g_n \equiv 1 - x_n - \sigma_w^2 - c_n; \quad A_n \equiv \{m_n^2 \leq g_n\}; \quad V_n(m) \equiv 1 - x_n - \sigma_w^2 + \text{Max}\{0, g_n - m^2\}; \quad \text{and } u_0 \equiv 1 - \hat{x} - \sigma_w^2 + EV_1(m_1)$$

then for each $n=1,2,\dots$,

$$u_0 = c_n + \delta \gamma E[V_{n+1}(m_{n+1}) | A_1, \dots, A_n] \quad (\text{no excess supply of young labor}); \quad (31a)$$

$$1 - \sigma_w^2 - x_n - c_n \geq 0 \quad (\text{Hiring occurs on grade } n \text{ for some } m_n \text{ values}). \quad (31b)$$

In the above $\{m_n\}_{n=1}^{\infty}$ is the stochastic process of means described in section 4.

It is easy to show that in any equilibrium the value functions $\{V_n\}_{n=1}^{\infty}$ obey the functional equation in (21) and (22). Given an equilibrium, using the notation in the definition above we see that the hiring rule is given by $\bar{m}_n \equiv g_n^{1/2}$, and the hiring set is given by A_n , for each n .

In appendix A we show that equilibria exist. Not only that, there are many equilibria. The equilibria are indexed by the value of the age one value function at $m=0$, $V_1(0)$. In particular, for each real number v greater than or equal to some critical value, which we refer to as \bar{v} , there exists an equilibrium with $V_1(0) = v$. (See remark R1 of Appendix A.)

The unique equilibrium such that $V_1(0) = \bar{v}$ will be referred to as the **bounded equilibrium**. (See Remark R2 of appendix A.) The bounded equilibrium has the following interpretation: Let us consider a finite horizon equilibrium where we insist that the decision problem is the same as our earlier description for all agents with technology of age $n \leq N+1$, but where we insist that no hiring is allowed on technologies of age $N+1$ or larger. This means that we "kill" all technologies precisely when they get to age $N+2$, regardless of whether or not young agents would want to work on such technologies. It turns out that there exists a unique finite horizon equilibrium of this sort. The unique bounded equilibrium is the limit of the finite horizon equilibria, as the horizon, N , goes to infinity.

Each equilibrium with $V_1(0)$ strictly greater than \bar{v} will be such that

$$\lim_{n \rightarrow \infty} V_n(m) = \infty \text{ for all } m. \quad (32)$$

We refer to such equilibria as **Ponzi equilibria**.

One thing that comes out of our construction of equilibria is:

Proposition 5.1.: In any equilibrium, the hiring rule, \bar{m}_n , is strictly positive for each $n=1,2,\dots$; and the hiring set $A_n=[-\bar{m}_n, \bar{m}_n]$ is a non-empty non-degenerate interval for each n .

Corollary 5.2: In any equilibrium, for each $N < \infty$, there exists a strictly positive probability that the length of life of the technological line will exceed N .

Below we show that in any Ponzi equilibrium c_n , the wage payment of the old to the young for their labor tends to minus infinity so will eventually be negative. This means that the young pay the old. This is in return for the information that the old will transfer to the young.

Proposition 5.3. In any Ponzi equilibrium $\lim_{n \rightarrow \infty} c_n = -\infty$.

It should be clear what is going on in the Ponzi equilibria. Suppose that at date 0 a young agent starts a technological line. Any arbitrarily high value for the date one value function, V_1 , may be obtained by setting c_1 equal to some large negative number. The date one young are willing to pay this amount because they expect that when they get old the young then will pay them back an even higher amount. That even higher amount is justified by yet higher payments. Etc. Etc.

Regardless of whether we are in a bounded or Ponzi equilibrium, the lifetime payoff to any agent is equal to $u_0 \equiv 1 - \hat{x} - \sigma_w^2 + \delta \gamma EV_1(m_1)$. Hence despite the fact that along a given Ponzi equilibrium $V_n(m) \rightarrow \infty$, each agent receives the same finite discounted payoff of u_0 . The reason for this is simple. Even though $V_n(m) \rightarrow \infty$, so too does $c_n \rightarrow -\infty$. The payoff to the agent who is young at n is $c_n + \delta \gamma EV_{n+1}(m_{n+1}) = u_0 < \infty$. Implicit in all of the paper and indeed in the

construction of equilibria above is the following: We rule out equilibria of a particular type. We do not consider equilibria where the age-one value function, $V_1(m)$, depends upon the date. Suppose that from dates 1 through T technological line A is in used. At date T no young are hired on that line and it dies. The young form a new technological line B at that date. Let V_1^A and V_1^B denote the age-one value function corresponding to technological lines A and B. Since we have multiple equilibria it is in principle possible that at date T the young form a technological line with a different associated Ponzi equilibrium than that of line A. In that case $V_1^A \neq V_1^B$. This is a form of time-nonstationarity. If we allowed for this inequality we would then have to index our value functions by not only the age of the technology, but also by the calendar time. We rule out such time nonstationarities. Note however that this does not imply in our model that $V_n = V_{n+1}$.

6. The Planner's Problem

6.1. The planner wishes to maximize the discounted sum of utilities of all the generations starting at date zero. This means he will maximize discounted aggregate output. As in the case of equilibrium, the planner's problem can be recast in a stationary form, with a discount factor of $\gamma\delta$.

The only difference between the Planner's problem and the equilibrium solution is in the hiring sets. The planner chooses the hiring set to maximize aggregate output of all generations. The equilibrium hiring set is chosen to maximize the payoff of the foreman doing the hiring. In this section we shall compare the hiring sets in the two situations.

In the equilibrium solution, the discounted aggregate output may be computed by adding the output of each foreman to that of the lifetime output of the current and future young. The young at any future date will have an expected output equal to u_0 given in (19). The equilibrium utility of an old agent with technology of grade n and beliefs represented by (m, x) was denoted by $V(m, x)$. The equilibrium lifetime utility of the young agent is u_0 . Hence the equilibrium utility of the foreman of type (m, x) plus the discounted lifetime utility of the young agent that he interviews, plus the discounted output of the agent that this young agent interviews in the subsequent period, and

so on into the infinite future, is equal to

$$W_n^e(m) \equiv V_n(m) + u_0/(1-\delta\gamma). \quad (33)$$

The value of W_n^e does not represent the maximum possible discounted output stream. The planner will wish to follow a hiring policy that will, as a function of m and x , maximize the analog of W_n^e .

The planner has the same information as the foreman does; that is, he knows m and x . If a foreman of type (m, x) hires a young agent, the planner's expectation of their joint output is

$$[1-x-\sigma_w^2] + [1-m^2-x-\sigma_w^2] = 2(1-x-\sigma_w^2) - m^2. \quad (34)$$

If the young agent is to work alone, the planner must assign him a random new technological line, and so he expects the joint output of the two agents to be

$$[1-x-\sigma_w^2] + [1-x-\sigma_w^2]. \quad (35)$$

Let $W_n(m)$ be the planner's expectation of the infinite sequence of such pairs of outputs. It solves the equation

$$W_n(m) = \text{Max} \{ 2(1-x_n-\sigma_w^2) - m^2 + \delta\gamma E[W_{n+1}(m_{n+1})|m_n=m], \\ 1-x_n-\sigma_w^2 + 1-x-\sigma_w^2 + \delta\gamma E W_1(m_1) \} \quad (36)$$

where the first expectation is over m_{n+1} given m_n (defined via (26) and (27)) and the second is over m_1 defined in (18). Using standard contraction mapping techniques it should be easy to see that there exists a unique sequence of continuous functions $\{W_n\}_{n=1}^\infty$ satisfying (36).

Lemma 6.1: $W_n(m)$ is decreasing in $|m|$ for each fixed n .

Since the top line of (36) is a decreasing function of $|m_n|$, or equivalently, of m^2 , while the second line of (36) does not depend on m_n , the planner will choose a cutoff value for $|m_n|$, call it m_n^{opt} and reject all matches for which $|m_n|$ exceeds this value:

$$m_n^{opt} = \{\hat{x} - x_n + \delta \gamma \{E[W(m_{n+1})|m_n^{opt}] - EW_1(m_1)\}\}^{1/2}. \quad (37)$$

The equilibrium hiring set cutoff, \bar{m}_n , is given by (23):

$$\bar{m}_n = \{\hat{x} - x_n + \delta \gamma \{E_2[V_{n+1}(m_{n+1})] - E_2 V_1(m_1)\}\}^{1/2}. \quad (38)$$

From (33) this becomes

$$\bar{m}_n = \{\hat{x} - x_n + \delta \gamma \{E_2[W_{n+1}^*(m_{n+1})] - E_2 W_1^*(m_1)\}\}^{1/2}. \quad (39)$$

Since W_n is the outcome of socially optimal decisions,

$$W_n(m) \geq W_n^*(m) \quad \text{for all } m. \quad (40)$$

Recall that E_2 in (38) represents expectations over m_{n+1} conditional only on information that the m_t fell in the hiring set, A_t , in each of the previous periods in the life of the current technology line.

Consider a planner at some fixed date t observing the equilibrium in process. At that date the old will be using a technology of age n . Suppose that the planner knows the value of m_n of

the old agent using that technology. If the planner had the opportunity to change the hiring decision at that date only, would the planner choose to increase or to decrease the hiring set? It turns out that the planner would want to decrease it!

Let m_n^* be the optimal hiring rule under the conditions just mentioned. The planner's value function is then given by the relation analogous to (36):

$$W_n(m) = \text{Max} \{ 2(1-x_n-\sigma_w^2) - m^2 + \delta \gamma E[W_{n+1}^e(m_{n+1})|m], \\ 1-x_n-\sigma_w^2 + 1 - \hat{x} - \sigma_w^2 + \delta \gamma E W_1^*(m_1) \} \quad (41)$$

Eq. (41) emphasizes the problem of the planner in our experiment: The planner is deciding the decision rule for a technology of age n at some fixed date t . Given m the planner must decide whether or not to allow a young to be matched with a foreman with mean m , or whether to inform the young to go it alone. If the young is hired at date t , then in each subsequent date the equilibrium is played out. Hence we see the equilibrium value function W_{n+1}^e on the top line of (41). Suppose on the other hand the planner recommends that no young are hired by the foreman of type m . Then the young goes it alone. If we suppose that young agents re-start the equilibrium outcome, then W_1^* on the second line of (41) is the equilibrium infinite horizon value function; i.e. $W_1^* = W_1^e$. If alternatively the young start-up the planner's problem, then $W_1^* = W_1$. From (40) we therefore have that in either case

$$W_1^*(m) \leq W_1(m) \text{ for all } m. \quad (42)$$

Using Corollary 4.2 in addition to arguments similar to that of Lemma 6.1, it is easy to see that the top line of (41) is decreasing in $|m|$. The second line in (41) is independent of m . Hence the solution to the problem in (41) and (42) is characterized by a hiring rule $m_n^* \geq 0$. The planner will

recommend that the young be hired whenever m_n lies in $[-m_n^*, m_n^*]$, and will recommend no hiring when m_n lies outside of this interval.

We now have the following:

Proposition 6.2: Suppose that the planner may make a one time change in the hiring rule. In particular suppose that the planner is solving the problem characterized by (41) and (42) for some fixed n at some date t . Then the planner will choose to reduce the hiring set; i.e., $m_n^* \leq \bar{m}_n$. The inequality is strict whenever $\bar{m}_n > 0$. This means that if there is ever a non-trivial amount of hiring, that amount hiring is excessive.

[Figure 3 about here].

Proposition 6.2 is true because the informational asymmetry between foremen and young agents gives rise to an adverse selection problem. The marginal technology, \bar{m}_n , is the worst among continuing technologies. When he hires, the marginal foreman ignores the fact that he is contaminating the distribution of surviving technologies. Therefore, in equilibrium, there is a tendency to stick too much to old methods and the startup rate of new technologies is too low.

6.2. Implementing the optimal policy. The planner can implement the optimal limited intervention policy without knowing any of the m 's. If he chooses to intervene at the point at which technologies are n periods old, equation (38) implies that he can induce old agents to be indifferent between hiring at $m = m_n^*$ by imposing a hiring tax τ_n given by

$$\tau_n = (\bar{m}_n)^2 - (m_n^*)^2. \quad (43)$$

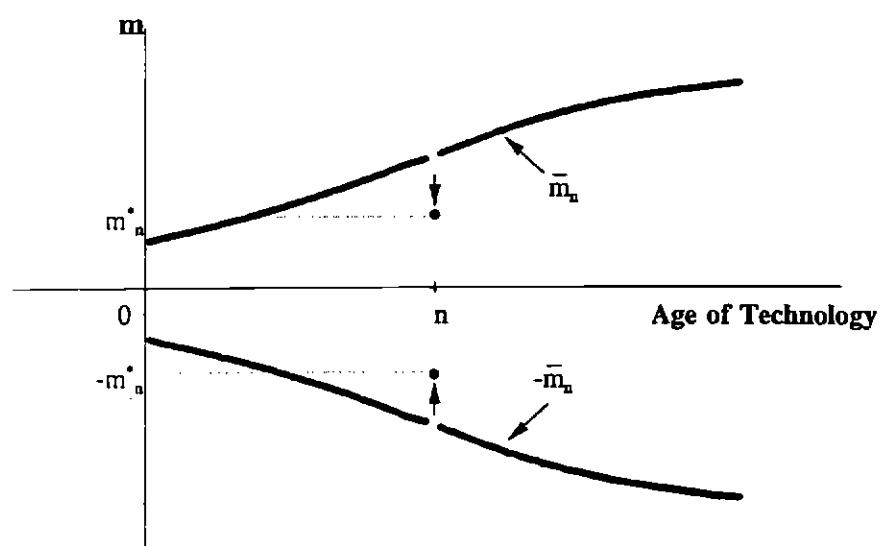


Figure 3: The Optimal Limited Intervention.

7. Overlapping Generations Learning Versus Frictionless Learning. How well does the OG mechanism approximate learning by, say, an infinitely-lived individual who sees the same evidence as the old and young agent combined? Since $h_1(x) = x\sigma_w^2/(x + \sigma_w^2)$, proposition 4.3 implies

$$\frac{\alpha x_n^2}{x_n + \sigma_w^2} \leq \sigma_{n+1}^2 \leq \frac{\alpha x_n^2}{x_n + \sigma_w^2} + \alpha \sigma_n^2. \quad (44)$$

We call this the overlapping generations (OG) case.

Alternatively, suppose that an infinitely lived individual sees the signals of both the young and old agents in each period. In this case, the difference equation for σ_n^2 becomes

$$\sigma_{n+1}^2 = \frac{\alpha x_n^2}{2x_n + \sigma_w^2} + \alpha \sigma_n^2, \quad (45)$$

since there are now two signals per period, and there is no truncation (i.e., $\bar{m} = \infty$). We call this the frictionless case.

The situation is shown in figure 4. The top two lines are straight, with slope α . The bottom line has slope zero. As x_n declines, all three lines in figure 4 shift down, but if all are evaluated at the same x_n , their relative positions remain unchanged. Equation (44) implies that for σ_n^2 given, σ_{n+1}^2 must, in the OG case, lie in the shaded area. So from any fixed prior variance x_n , the frictionless case involves faster learning, which is reflected in the greater height of its curve, and a more rapid growth in σ_n^2 . The least learning (none!) occurs for the case $\alpha = 0$, in which case $\sigma_{n+1}^2 = 0$ for all n , for both the OG and the frictionless cases.

[Figure 4 about here].

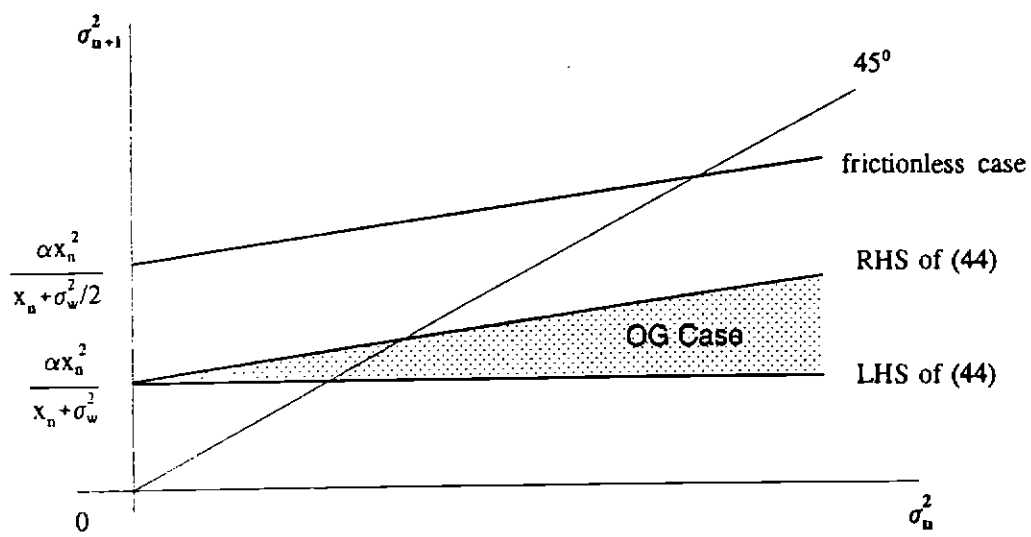


Figure 4: The σ_n^2 process in the OG and the frictionless cases.

The faster learning in the frictionless case also implies a lower x_n sequence. From equation (10), the OG case implies the following difference equation for x_n on a surviving technology:

$$x_{n+1} = \frac{\alpha x_n^2}{x_n + \sigma_w^2} + \sigma_\epsilon^2,$$

whereas in the frictionless case it is

$$x_{n+1} = \frac{\alpha x_n^2}{2x_n + \sigma_w^2} + \sigma_\epsilon^2.$$

These two curves are drawn in figure 5; the top one pertains to the OG case. Taking \hat{x} as the initial condition for both cases implies that at each date x_n is higher for the OG case, and it must also have a higher limit, denoted by \bar{x}_{OG} as opposed to \bar{x}_F for the frictionless case. Output and productivity are, at each date, both lower in the OG case, not just at each date, but in the limit as well.

[Figure 5 around here].

If there were no upgrading, the productivity differences between the OG and the frictionless case would disappear in the long run, because even OG learning would eventually perfectly reveal the parameter(s) of interest. But because of continual upgrading, ignorance about θ_a is always there, but it is permanently larger in the OG case. In the limit, both economies grow at the rate $\gamma-1$, but the differences in productivity levels remain.

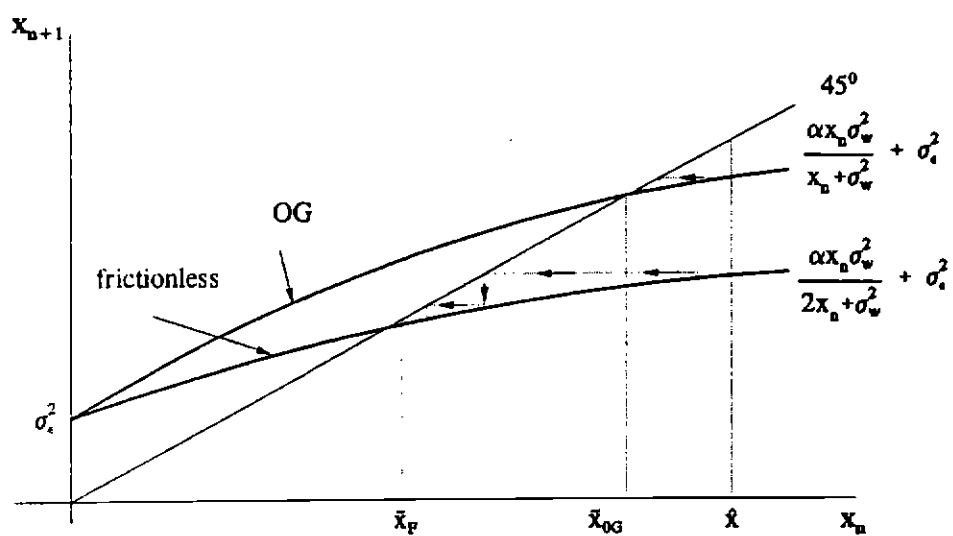


Figure 5: Evolution of x_n in the two cases.

8. Productivity, Failure, and the Earning-Experience Profile

8.1. Productivity and Failure If an old agent does not hire anyone, his technological line will end one period later. We define "failure" to be the cessation of hiring. In this model, less productive technologies are the more likely to fail. When $|m| < \bar{m}$, output per man of the two-agent team is

$$\frac{1}{2}(1 - x - \sigma_w^2 + 1 - x - \sigma_w^2 - m^2) = 1 - x - \sigma_w^2 - \frac{m^2}{2} \equiv q_n(m) .$$

The probability that the firm will fail next period is

$$Q_n(m) = \Phi\left(\frac{-\bar{m}_{n+1} - \alpha^{1/2}m}{\alpha^{1/2}\sqrt{h_1(x_n)x_n/\sigma_w^2}}\right) + 1 - \Phi\left(\frac{\bar{m}_{n+1} - \alpha^{1/2}m}{\alpha^{1/2}\sqrt{h_1(x_n)x_n/\sigma_w^2}}\right)$$

where Φ is the normal cumulative distribution function. The function $Q_n(m)$ is increasing in $|m|$. The relevant values of $|m|$ are those below \bar{m} , and here q_n and Q_n are negatively related as $|m|$ varies. This is shown in figure 6. This implication fits the stylized fact that less productive firms are more likely to fail. The economics is that such firms fail because they find themselves in technologies that young agents find hard to assimilate; this may well be an important cause of failure in high-tech industries.

[figure 6 about here].

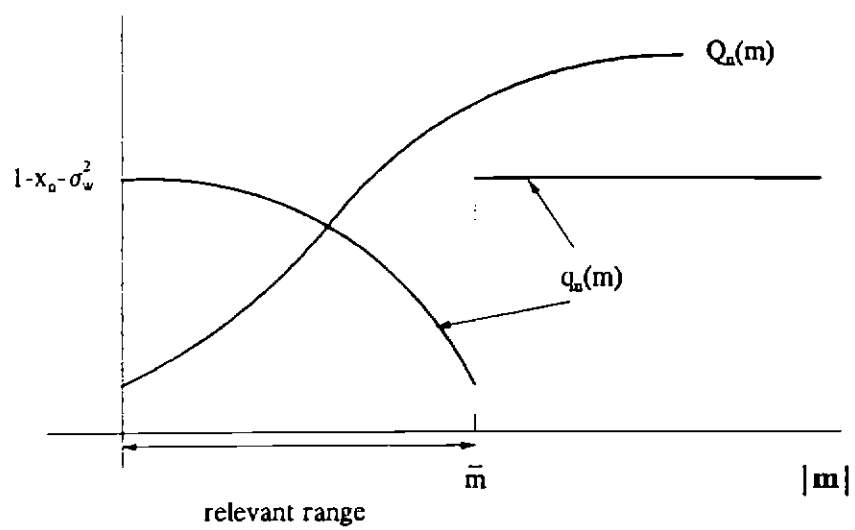


Figure 6: The negative relation between q_n and Q_n .

8.2. Human Capital and the Earnings - Experience Profile. Old agents who operate older technological lines have higher productivity because they have a lower $x_n = B^n(\hat{x})$. When hired to work on an age- n technology, a young agent receives a level of consumption that (in view of (17) and (24)) equals

$$c_n = u_0 - \gamma \delta E_n V_{n+1}(m)$$

When old, this agent's expected consumption will be $E_n V_{n+1}(m)$. Hence the age-expected earnings profile of this agent will have slope

$$\frac{E_n V_{n+1}(m)}{c_n} = \left[\frac{u_0}{E_n V_{n+1}(m)} - \gamma \delta \right]^{-1}$$

On the right-hand side, only V_{n+1} depends on n , and so the slope of the age-earnings profile will be relatively high in an age- n technology if its V_{n+1} is relatively high.

9. Conclusion

We have modelled training, learning by doing, and production as joint activities. Training, however, occurs only if a young agent works together with an older one, and it consists of the younger agent watching the older one do the job. In equilibrium, some young agents start new firms on their own, while others join an older agent as apprentices. Apprentices give up part of their output in return for the training that they get from older agents, training that will raise their second period productivity.

The overlapping generations model differs radically from the single agent infinite horizon model studied in Jovanovic and Nyarko (1994). First, there is an avoidable adjustment cost -- modelled as waiting time -- in the process by which knowledge is transferred from the old to the

young. This leads to a permanently lower productivity in the overlapping generations model. Second, the non-Ponzi equilibrium is constrained inefficient, in that too few new startups take place, and too many old technologies are kept alive. This is because the informational asymmetry between the old and the young gives rise to an adverse selection problem: an older agent with inferior skills pays his apprentice a wage that reflects not his own skills, but rather the skills of the average old agent.

10. Appendix A: Construction and Properties of Equilibria.

In the construction below, the reader may find figure 7 useful. First, we provide a summary of what is to come. We first define by induction a sequence of non-negative numbers $\{\Gamma_n\}_{n=1}^{\infty}$ and a sequence of functions $G_n: [\Gamma_{n-1}, \infty) \rightarrow \mathbb{R}$ for $n=1, 2, \dots$. We will then define for each $g_1 > 0$ sufficiently large, the sequence of numbers $\{g_n\}_{n=1}^{\infty}$ by setting $g_n = G_n(g_1)$. This sequence of numbers will then be shown to define an equilibrium. In particular, we obtain a multiplicity of equilibria with each equilibrium indexed by g_1 . Now the details!

For each $g_1 > 0$ define

$$U_1(g_1) = \delta \gamma E_1[\text{Max}\{0, g_1 - m_1^2\}] \quad (46)$$

where the expectation E_1 is over the distribution of m_1 . Define for each $n=1, 2, \dots$,

$$h_n = \hat{x} - x_n + \delta \gamma \{\hat{x} - x_{n+1}\} \text{ and} \quad (47)$$

$$H_n(g_1) = h_n - U_1(g_1). \quad (48)$$

Note that for each n , $H_n(0) > 0$, $H_n(g_1)$ is strictly decreasing in g_1 and $H_n(g_1) \rightarrow -\infty$ as $g_1 \rightarrow \infty$.

Define $\Gamma_0 \equiv 0$ and $G_1(g_1) \equiv g_1$ on $[\Gamma_0, \infty)$. We proceed by induction. Suppose that for some $N \geq 1$ we have defined the non-negative numbers $\Gamma_0, \Gamma_1, \dots, \Gamma_{N-1}$ and the functions $G_n: [\Gamma_{n-1}, \infty) \rightarrow \mathbb{R}$ for

$n=1, \dots, N$, with $G_n(g_1)$ strictly increasing, taking the value 0 at $g_1 = \Gamma_{n-1}$ and ∞ at $g_1 = \infty$. Note that we have already shown this for $N=1$. Define $\Gamma_N > 0$ be the unique number such that $G_N(g_1) - H_N(g_1) = 0$ at $g_1 = \Gamma_N$ and $G_N(g_1) - H_N(g_1) > 0$ at each $g_1 > \Gamma_N$. From the properties of H_N and G_N such a $\Gamma_N > 0$ exists. Define $U_{N+1}: \mathbb{R}_+^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$ by

$$U_{N+1}(g_{N+1}, g_1) = \delta \gamma E_{N+1}[\text{Max } \{0, g_{N+1} - m_{N+1}^2\} | A_1, \dots, A_N] \quad (49)$$

where the expectation above is with respect to the distribution of m_{N+1} conditional on the events $A_n = \{m_n \in [-(G_n(g_1))^{1/2}, (G_n(g_1))^{1/2}]\}$ for $n=1, \dots, N$. For fixed $g_1 \geq 0$, the function $U_{N+1}(\cdot, g_1)$ of g_{N+1} is zero at $g_{N+1} = 0$, and is strictly increasing taking the value ∞ at $g_{N+1} = \infty$. Define the function $G_{N+1}: [\Gamma_N, \infty) \rightarrow \mathbb{R}$ as follows: for each $g_1 \geq \Gamma_N$ let $G_{N+1}(g_1)$ be the unique value of g_{N+1} such that

$$U_{N+1}(g_{N+1}, g_1) = G_N(g_1) - H_N(g_1). \quad (50)$$

From the properties of the U_{N+1} and $G_N(g_1) - H_N(g_1)$ just mentioned, it should be easy to see that the function $G_{N+1}(g_1)$ is well-defined on $[\Gamma_N, \infty)$, and is strictly increasing taking the value 0 at $g_1 = \Gamma_{N+1}$ and ∞ at $g_1 = \infty$.

Hence by induction we have constructed a sequence of numbers $\{\Gamma_n\}_{n=0}^\infty$ and a sequence of functions $\{G_n\}_{n=1}^\infty$ with $G_n: [\Gamma_{n-1}, \infty) \rightarrow \mathbb{R}$ strictly increasing from 0 to ∞ . Since by construction Γ_n is the value of g_1 where $H_n(g_1) = G_n(g_1)$. As n gets larger it should be clear (see fig. 7) that the functions H_n and G_n both shift to the right. Hence $\Gamma_{n+1} < \Gamma_n$ for all $n \geq 1$. Since $H_n(g_1)$ increases monotonically to $H_\infty(g_1) \equiv \lim_{n \rightarrow \infty} H_n(g_1)$, it should be clear that $\Gamma_n \leq H_\infty^{-1}(0)$ for all n . Hence $\lim_{n \rightarrow \infty} \Gamma_n = \Gamma_\infty$ exists and is finite. Each of the functions $G_n(g_1)$ for $n \geq 1$ is therefore well-defined on $[\Gamma_\infty, \infty)$.

[Figure 7 about here].

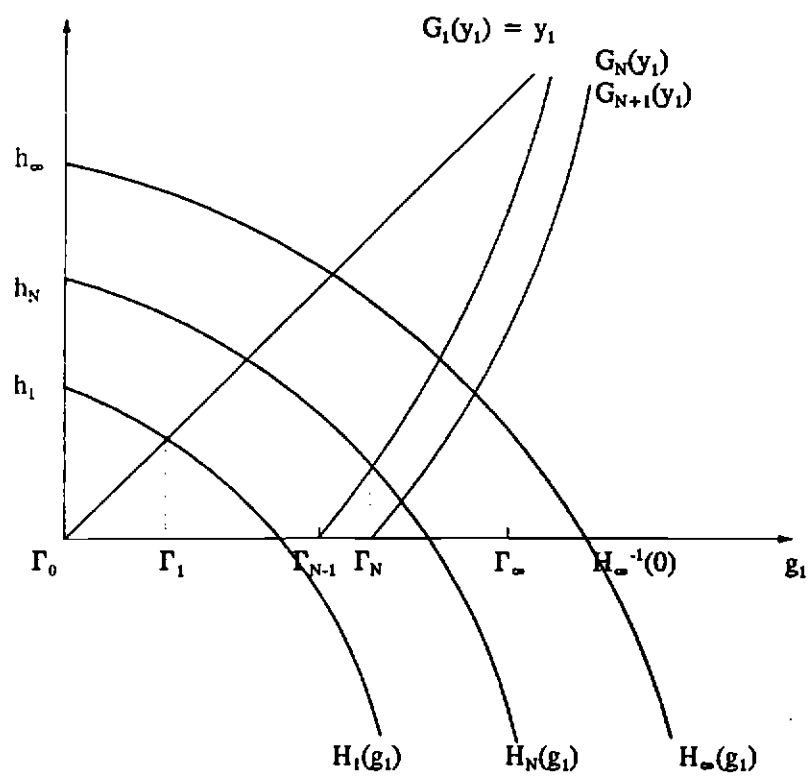


Figure 7: Construction of equilibrium.

Fix any $g_1 \geq \Gamma_\infty$. Define for each $n > 0$, $g_n = G_n(g_1)$ and $\bar{m}_n = g_n^{1/2}$. Define

$$V_n(m) = 1 - \sigma_w^2 - x_n + \text{Max} \{0, g_n - m^2\} \quad \text{where } x_n = B^n(x_0). \quad (51)$$

Equating (49) and (50) results in

$$\begin{aligned} g_N &= H_N(g_1) + U_{N+1}(g_{N+1}, g_1) \\ &= \hat{x} - x_N + \delta \gamma \{ \hat{x} - x_{N+1} + E_{N+1}[\text{Max} \{0, g_{N+1} - m_{N+1}^2\} | A_1, \dots, A_N] - E_1[\text{Max} \{0, g_1 - m_1^2\}] \} \\ &= \hat{x} - x_N + \delta \gamma \{ [1 - \sigma_w^2 - x_{N+1} + E_{N+1}[\text{Max} \{0, g_{N+1} - m_{N+1}^2\} | A_1, \dots, A_N]] - [1 - \sigma_w^2 - \hat{x} + E_1[\text{Max} \{0, g_1 - m_1^2\}]] \} \\ &= \hat{x} - x_N + \delta \gamma \{ E_{N+1}[V_{N+1}(m_{N+1}) | A_1, \dots, A_N] - E_1[V_1(m_1)] \}. \end{aligned}$$

Putting this in (51) gives

$$V_N(m) = 1 - \sigma_w^2 - x_N + \text{Max} \{0, 1 - \sigma_w^2 - x_N - m^2 - c_n\} \quad \text{where}$$

$$c_n \equiv 1 - \hat{x} - \sigma_w^2 - \delta \gamma \{ E_{N+1}[V_{N+1}(m_{N+1}) | A_1, \dots, A_N] - E_1[V_1(m_1)] \}.$$

Since we may write c_n above as

$$c_n = u_0 - \delta \gamma \{ E_{N+1}[V_{N+1}(m_{N+1}) | A_1, \dots, A_N] \} \quad \text{where } u_0 = 1 - \hat{x} - \sigma_w^2 + \delta \gamma E_1[V_1(m_1)]$$

it should be clear that our construction satisfies the definition of an equilibrium. Hence each $g_1 \geq \Gamma_\infty$ defines an equilibrium. ■

Remarks:

R1: (Multiplicity of Equilibria) From the above construction we have a multiplicity of equilibria indexed by g_1 in the set $[\Gamma_\infty, \infty)$. Note that g_1 and $V_1(0)$ are related by the relation $V_1(0) = 1 - x_1 - \sigma_w^2 + g_1$. When $g_1 = \Gamma_\infty$, $V_1(0) = 1 - x_1 - \sigma_w^2 + \Gamma_\infty \equiv \bar{v}$, say. Hence we could equivalently index the equilibria by the value of $V_1(0)$ so long as it lies in the set $[\bar{v}, \infty)$.

R2: (The Unique Bounded Infinite-Horizon Equilibrium): The equilibrium from $g_1 = \Gamma_\infty$ is such that if V_n denotes the associated value functions then $\lim_{n \rightarrow \infty} V_n(m) \leq K$ for all m , where K is a constant independent of n . (Indeed $K = 1 - x_\infty - \sigma_w^2 + h_w / (1 - \delta \gamma)$.) This is the content of Proposition 10.1 below. We mention in the remark R3 below that all equilibria from $g_1 > \Gamma_\infty$ will have $\lim_{n \rightarrow \infty} V_n(m) = \infty$ for all m . Hence the equilibrium associated with $g_1 = \Gamma_\infty$ is therefore referred to as the *bounded* equilibrium. Proposition 10.1 obtains the bounded equilibrium via a construction which enables us to interpret it as the limit as $N \rightarrow \infty$ of N -horizon equilibria. The latter are equilibria where we insist that the decision problem is the same as the earlier analysis for all agents with technology of age $n \leq N+1$, but where we insist that no hiring is allowed on technologies of age $N+1$ or larger. This equilibrium is obtained by setting $g_1 = \Gamma_n$ and $g_n = G_n(g_1)$ for $n = 1, 2, \dots, N+1$. (In this case $g_{N+1} = 0$ and $G_r(g_1)$ is not well-defined for $r > N+1$.) There is a unique N -horizon equilibrium of this type. The unique bounded equilibrium is the limit of the finite horizon equilibria as the horizon, N , goes to infinity.

R3 (The Ponzi Equilibria): In proposition 10.2 below we show that for each equilibrium corresponding to a g_1 strictly greater than Γ_∞ will have an unbounded limiting value function, i.e., $\lim_{n \rightarrow \infty} V_n(m) = \infty$ for all n . We refer to these equilibria as Ponzi equilibria.

R4 (Ruling out Equilibria where there is never hiring on some grade):

Suppose that in an equilibrium there is some hiring on grade n but no hiring on grade $n+1$. Hiring on grade n (by the foreman of type $m_n = 0$) implies that $1 - x_n - \sigma_w^2 \geq c_n$ (i.e., firms want to hire) and $u_0 = c_n + \delta \gamma E_n V_{n+1}(m_{n+1}) = c_n + \delta \gamma [1 - x_{n+1} - \sigma_w^2]$ (i.e., the young want to be hired) which implies that

$$u_0 - \delta \gamma [1 - x_{n+1} - \sigma_w^2] \leq 1 - x_n - \sigma_w^2. \quad (52)$$

For there to be no hiring on grade $n+1$, $1 - x_{n+1} - \sigma_w^2 \leq c_{n+1}$ and $u_0 \geq c_{n+1} + \delta \gamma [1 - x_{n+2} - \sigma_w^2]$ which implies that $1 - x_{n+1} - \sigma_w^2 \leq u_0 - \delta \gamma [1 - x_{n+2} - \sigma_w^2]$. However since $x_{n+2} < x_{n+1} < x_n$, this is easily seen to be a contradiction to (52). Hence if there is hiring on grade n there is also hiring on grade $n+1$.

It therefore remains only to show that there is hiring on grade 1. Suppose, per absurdum there is no hiring on grade 1. Then it is easy to see that (52) holds for $n=0$ if we define $x_0 = \hat{x}$. Mimicking the previous argument shows that assuming there is no hiring at date 1 leads to a contradiction to equation (52) with $n=0$. Hence in any equilibrium there is hiring on each grade. ■

Proposition 10.1. For all $n=1,2,\dots$, $G_n(\Gamma_\infty) < h_\infty/(1-\delta\gamma) < \infty$. In particular, if $\{\bar{m}_t, V_t\}_{t=1}^\infty$ is the equilibrium associated with $g_1 = \Gamma_\infty$ as constructed above, then for each m , $\lim_{t \rightarrow \infty} V_t(m) \leq 1 - x_n - \sigma_w^2 + h_\infty/(1-\delta\gamma) < \infty$.

Proof of Proposition 10.1: For each n , it is easy to see that $G_n(g_1)$ is continuous on $[\Gamma_1, \infty)$. We will later on show that

$$G_n(\Gamma_N) \leq h_\infty/(1-\delta\gamma) \text{ for each } n \text{ and } N \text{ with } N > n. \quad (53)$$

The first part of this proposition therefore follows immediately from taking limits as $N \rightarrow \infty$ in (53). The second part of the proposition follows immediately from the first part and the definition of $V_n(m)$ in (51).

We now prove (53). Fix any $N < \infty$. Define $g_1 = \Gamma_N$ and $g_n = G_n(g_1)$ for all $n=1,2,\dots,N+1$. From the earlier construction we know that such g_1, \dots, g_{N+1} is well-defined with $g_{N+1} = 0$. Further,

$$g_n = U_{n+1}(g_{n+1}, g_1) + H_n(g_1) \text{ for each } n=1,\dots,N. \quad (54)$$

However, from the definition of U_{n+1} in (49), $U_{n+1}(g_{n+1}, g_1) \leq \delta \gamma g_{n+1}$. Also from (48), $H_n(g_1) \leq h_\infty$. Hence (54) becomes, $g_n \leq \delta \gamma g_{n+1} + h_\infty$ for all $n=1, \dots, N$. This is equivalent to

$$g_{N-r} \leq \delta \gamma g_{N-(r-1)} + h_\infty \quad \text{for all } r=0, 1, \dots, N-1. \quad (55)$$

Define $q_r \equiv g_{N-r+1}$ for all $r=0, 1, \dots, N-1$. Then (55) becomes

$$q_{r+1} \leq \delta \gamma q_r + h_\infty \quad \text{for all } r=0, 1, \dots, N-1. \quad (56)$$

Note that $q_0 = g_{N+1} = 0$. Define the linear function $f: \mathbb{R} \rightarrow \mathbb{R}$ by $f(q) \equiv (\delta \gamma)q + h_\infty$. Iterates under this map from the origin will always remain below the fixed point, $h_\infty/(1-\delta \gamma)$, of this mapping. Since the right hand side of (56) is $f(q_r)$, we conclude that $q_r \leq h_\infty/(1-\delta \gamma)$ for all $r=0, 1, \dots, N-1$. This in turn implies (53). ■

Proposition 10.2.: Fix any g_1 strictly greater than Γ_∞ . Then $\lim_{n \rightarrow \infty} G_n(g_1) = \infty$. In particular, if $\{\bar{m}_i, V_i\}_{i=1}^\infty$ is the equilibrium associated with any $g_1 > \Gamma_\infty$ as constructed above, then for each m , $\lim_{i \rightarrow \infty} V_i(m) = \infty$.

Proof of Proposition 10.2.: The second part of the proposition follows immediately from the first part and the definition of $V_n(m)$ in (51). We now prove the first part of the proposition. We will later on show that

$$dG_n(g_1)/dg_1 \geq 1/(\delta \gamma)^n \quad \text{for all } n \geq 1 \quad \text{and at each } g_1 \geq \Gamma_{n-1}. \quad (57)$$

Fix any $g_1 > \Gamma_\infty$. (57) implies that for all n , $G_n(g_1) - G_n(\Gamma_\infty) \geq (g_1 - \Gamma_\infty)/(\delta \gamma)^n$. Taking limits as $n \rightarrow \infty$ therefore implies that $\lim_{n \rightarrow \infty} G_n(g_1) = \infty$. Since $V_n(m) = 1 - \sigma_w^2 x_n + \text{Max}\{0, g_n - m^2\}$, this implies that $V_n(m) \rightarrow \infty$.

We now prove (57) by induction. Since $G_1(g_1) = g_1$ we see that the claim holds for $n=1$. So

assume that (57) is true for some n . Putting $g_n = G_n(g_1)$ in (50) implies that

$$U_{n+1}(G_{n+1}(g_1), g_1) = G_n(g_1) - H_n(g_1). \quad (58)$$

Totally differentiating this with respect to g_1 and using the fact that H_n is decreasing in g_1 implies that

$$(dU_{n+1}/dg_{n+1})(dG_{n+1}/dg_1) + (dU_{n+1}/dg_1) = dG_n/dg_1 - dH_n/dg_1. \quad (59)$$

Using the induction hypothesis and the fact that $dH_n/dg_1 < 0$ we conclude that

$$(dU_{n+1}/dg_{n+1})(dG_{n+1}/dg_1) + (dU_{n+1}/dg_1) \geq 1/(\delta\gamma)^n. \quad (60)$$

We will later show that for all n ,

$$0 < dU_{n+1}/dg_{n+1} \leq \delta\gamma \quad \text{and} \quad dU_{n+1}/dg_1 < 0. \quad (61)$$

Putting (61) into (60) implies that $dG_{n+1}/dg_1 \geq 1/(\delta\gamma)^{n+1}$ which is the induction step for $n+1$. Hence by induction this shows that (57) is true for all n .

We now prove (61). Fix any $n \geq 1$. It should be clear from the definition of U_n in (49) that for fixed g_1 , an increase in g_n by one unit will increase U_n , but by no more than $\delta\gamma$ units. Hence $0 < dU_{n+1}/dg_{n+1} \leq 1/\delta\gamma$. This proves the first part of (61). Now let us consider the effect of an increase in g_1 keeping g_{n+1} fixed. For any $r=1, \dots, n$, since $G_r(g_1)$ is increasing in g_1 , an increase in g_1 has the effect of increasing the conditioning sets A_r . It is easily shown that this implies that the variance of m_n will increase, with its mean remaining equal to zero. This in turn will lead to a reduction in the value of the integral in (49) which defines U_n . In particular, an increase in g_1 keeping g_{n+1} fixed will result in a decrease in U_{n+1} , so $dU_n/dg_1 < 0$. This proves the second part of (61). ■

11. Appendix B: Justification for Assumption (A.2). One implication of this assumption is that each continuing technological line will be upgraded in every period. In this appendix, we show that such upgrading will be myopically optimal for some parameter values. Since old agents are by assumption myopic, the term "myopically" refers to the young agents. In this sense, myopically optimal policies maximize the expected output of the old and the young combined, but they ignore the future value of the training that takes place during the period.

Assume that it is the old agent who decides on whether to switch or not switch, in addition to deciding whether to hire or not hire. Then we have the matrix of the total expected outputs of the two agents:²

	Don't Hire	Hire
No Switch	$1 - x - \sigma_w^2 + \gamma[1 - \hat{x} - \sigma_w^2]$	$1 - x - \sigma_w^2 + 1 - x - m^2 - \sigma_w^2 + 2(1 - x - \sigma_w^2) - m^2$
Switch	$\gamma[1 - \alpha x - \sigma_e^2 - \sigma_w^2 + 1 - \hat{x} - \sigma_w^2]$	$\gamma[1 - \alpha x - \sigma_e^2 - \sigma_w^2 + 1 - \alpha x - \sigma_e^2 - m^2 - \sigma_w^2]$ $+ \gamma[2(1 - \alpha x - \sigma_e^2 - \sigma_w^2) - m^2]$

If he doesn't hire, the old agent will prefer to switch if

$$1 - x - \sigma_w^2 < \gamma[1 - \alpha x - \sigma_e^2 - \sigma_w^2] \quad (62)$$

If he does hire, he will prefer to switch if

² If the young agent goes it alone, it is always optimal for him to start out on the latest available technology. This is because $\hat{x} = \alpha\hat{x} + \sigma_e^2$ so that $\gamma[1 - \alpha x - \sigma_e^2 - \sigma_w^2]$ always exceeds $1 - x - \sigma_w^2$. This fact has been incorporated into the (1,1) box of the matrix.

$$1 - x - \sigma_w^2 < \gamma[1 - \alpha x - \sigma_\epsilon^2 - \sigma_w^2] - \frac{(\gamma - 1)}{2} m^2 \quad (63)$$

Clearly, if (63) is met, then (62) is as well. If (63) holds for some m^2 , then it holds for all m^2 less than that value. The largest possible m^2 that, conditional on switching, will maintain the match is the one that equates the output under going it alone to the output under the match:

$$\begin{aligned} (\bar{m})^2 &= 2(1 - \alpha x - \sigma_\epsilon^2 - \sigma_w^2) - (1 - \alpha x - \sigma_\epsilon^2 - \sigma_w^2) - (1 - \hat{x} - \sigma_w^2) \\ &= \hat{x} - \alpha x - \sigma_\epsilon^2 \leq \hat{x} - \sigma_\epsilon^2 = \frac{\alpha \sigma_\epsilon^2}{1 - \alpha} \end{aligned}$$

Since $m^2 \leq (\bar{m})^2$, substitution for m^2 in (63) leads to the inequality

$$1 - x - \sigma_w^2 < \gamma(1 - \alpha x - \sigma_\epsilon^2 - \sigma_w^2) - \frac{(\gamma - 1)\alpha \sigma_\epsilon^2}{2(1 - \alpha)} \quad (64)$$

When $1 - \sigma_w^2 > 0$, the term $(1 - \alpha x - \sigma_\epsilon^2 - \sigma_w^2)$ is strictly positive for small enough α and σ_ϵ^2 . Then (64) will hold when α and σ_ϵ^2 are small enough and/or when γ is large enough. So we have proved:

Proposition 11.1: There exist parameter values under which it is myopically optimal for all agents to upgrade their technologies in each period.

12. Appendix C: The Proofs.

Proof of Proposition 4.1.: Let ϕ be any density function of some random variable η taking values on the real line. Suppose that ϕ is symmetric about the origin and suppose that $\phi(\eta)$ is strictly decreasing in $|\eta|$. Let Φ be the cumulative distribution function corresponding to ϕ . Fix any $k > 0$ and any $M \geq 0$. Then

$$\text{Prob} (\{(M+\eta)^2 \leq k\}) = \text{Prob} (\{-k+M \leq \eta \leq k+M\}) = \Phi(k+M) - \Phi(-k+M). \quad (65)$$

However, taking derivatives with respect to M in (65) yields

$$\partial/\partial M [\text{Prob} (\{(M+\eta)^2 \leq k\})] = \phi(k+M) - \phi(-k+M). \quad (66)$$

Under the assumptions on ϕ it is easy to see that for $k > 0$ and $M > 0$, $\phi(k+M) - \phi(-k+M) < 0$. Hence we conclude that $\text{Prob} (\{(M+\eta)^2 \leq k\})$ is decreasing in M .

Let η be normal with mean equal to zero and variance $\alpha h_1(x_n)x_n/\sigma_w^2$. From (65) and (66), m_{n+1} is a random variable whose distribution conditional on m_n is the same as $\alpha^{1/2}m_n + \eta$. Fix any $k > 0$. Then

$$\text{Prob} (\{m_{n+1}^2 \leq k | m_n\}) = \text{Prob} (\{(\alpha^{1/2}m_n + \eta)^2 \leq k\} | m_n). \quad (67)$$

Setting $M = \alpha^{1/2}m_n$ in our earlier argument shows that (67) is decreasing in m_n . This proves part (a) of our proposition.

(b) This follows immediately from part (a).

(c) Let $f: \mathbb{R}_+ \rightarrow \mathbb{R}$ be any strictly increasing function. Since conditional on m_n the distribution of m_{n+1} is independent of the acceptance sets, we may write

$$E_2[f(m_{n+1}^2)] = E_2[f(m_{n+1}^2)|A_1, \dots, A_n] = E_2[E[f(m_{n+1}^2)|m_n]|A_1, \dots, A_n], \quad (68)$$

where the outer expectation in the last term above is the expectation operator over m_n conditional on the events A_1, \dots, A_n . From part (a), the bigger is m_n the stochastically larger is the distribution of m_{n+1}^2 . That distribution is therefore largest at \bar{m}_n , the largest value of m_n in A_n . Hence from (68) we conclude that $E_2[f(m_{n+1}^2)] \leq E_2[f(m_{n+1}^2)|m_n = \bar{m}_n]$. Since f is an arbitrary strictly increasing function this implies the conclusion of the proposition as regards m_{n+1}^2 . Similar arguments prove the conclusion involving $|m_{n+1}|$. ■

Proof of Proposition 4.3: From (29)

$$\sigma_{n+1}^2 = \frac{\alpha h_1(x_n)x_n}{\sigma_w^2} + \alpha \hat{\sigma}_n^2$$

where $\hat{\sigma}_n^2$ is the post truncation variance of m_n . Since $\alpha \geq 0$, and $0 \leq \hat{\sigma}_n^2 \leq \sigma_n^2$, the

claim follows. ■

Proof of Proposition 5.3. This follows immediately from taking limits as $n \rightarrow \infty$ in (20), and using the defining property of a Ponzi equilibrium. ■

Proof of Lemma 6.1: Define $C(N \times \mathbb{R}) = \{W: N \times \mathbb{R} \rightarrow \mathbb{R} \text{ such that } W \text{ is continuous and bounded}\}$ and

$C_d(N \times \mathbb{R}) = \{W \text{ in } C(N \times \mathbb{R}) \text{ which are decreasing in } |m|\}$. Then we may consider (36) as defining for each W in $C(N \times \mathbb{R})$ a unique function $(TW): N \times \mathbb{R} \rightarrow \mathbb{R}$. It is easy to check that when W is continuous and bounded so is TW . Hence $T: C(N \times \mathbb{R}) \rightarrow C(N \times \mathbb{R})$. Further, T is a contraction operator. It therefore has a unique fixed point in $C(N \times \mathbb{R})$. That fixed point is the function $W(n, m) = W_n(m)$. From Corollary 4.2, it is easy to show that if W lies in $C_d(N \times \mathbb{R})$ then so too does TW . Hence $T: C_d(N \times \mathbb{R}) \rightarrow C_d(N \times \mathbb{R})$. The contraction mapping theorem therefore shows that $W_n(m)$ is decreasing in $|m|$. ■

Proof of Proposition 6.2: From (41) and (39),

$$(\delta \gamma)^{-1}[(m_n^{*2} - (\bar{m}_n)^2)] = \{E[W_{n+1}^e(m_{n+1})|\bar{m}_n^* - E_2[W_{n+1}^e(m_{n+1})]]\} - \{E_1 W_1^*(m_1) - E W_1^e(m_1)\}. \quad (69)$$

From (42), this becomes

$$(\delta \gamma)^{-1}[(m_n^{*2} - (\bar{m}_n)^2)] \leq E[W_{n+1}^e(m_{n+1})|\bar{m}_n^*] - E_2[W_{n+1}^e(m_{n+1})]. \quad (70)$$

If we let A_1, \dots, A_n denote the equilibrium hiring sets, then

$$E_2[W_{n+1}^e(m_{n+1})] = E_2[W_{n+1}^e(m_{n+1})|A_1, \dots, A_n] = E[E[W_{n+1}^e(m_{n+1})|m_n]|A_1, \dots, A_n]. \quad (71)$$

where expectations operator $E[.]|A_1, \dots, A_n$ is over m_n conditional on the hiring sets A_1, \dots, A_n , and $E[.]|m_n$ is that over m_{n+1} conditional on m_n . From Corollary 4.2, the distribution of $|m_{n+1}|$ conditional on m_n is stochastically increasing in the absolute value of m_n . From lemma 6.1, $W_{n+1}(m)$ is decreasing in $|m|$. Since $A_n = [-\bar{m}_n, \bar{m}_n]$, this implies that the right hand side of (71) is less than $E[W_{n+1}^e(m_{n+1})|\bar{m}_n]$. Putting this in (70) implies that

$$(\delta \gamma)^{-1}[(m_n^* - \bar{m}_n)^2] \leq E[W_{n+1}^e(m_{n+1})|m_n^*] - E[W_{n+1}^e(m_{n+1})|\bar{m}_n]. \quad (72)$$

Suppose per absurdum that $m_n^* > \bar{m}_n$. Then the left hand side of (72) is positive. Further, Corollary 4.2 implies that the right hand side of (72) is negative. This is a contradiction. Hence $m_n^* \leq \bar{m}_n$. ■

13. References

- Barron, John, Dan Black, and Mark Lowenstein, 1989, Job Matching and On-the-Job Training, *Journal of Labor Economics* 7, 1-29.
- Chari, V.V., and Hugo Hopenhayn, 1991, Vintage Human Capital, *Journal of Political Economy* 99, 1142-1165.
- Jovanovic, Boyan, and Yaw Nyarko, 1994, The Bayesian Foundations of Learning by Doing, unpublished paper, New York University.
- Prescott, Edward, and John Boyd, Dynamic Coalitions, 1987, Engines of Growth, *American Economic Review* 77, 63-67.